The State University of New York at Buffalo

CSE 574
Introduction to Machine Learning

Programming Assignment -1

Team Members:

Mohan Krishna Gaddam (mgaddam)
Jai Kishore Enamandala (Jaikisho)
Kartik Bapna (kartikba)

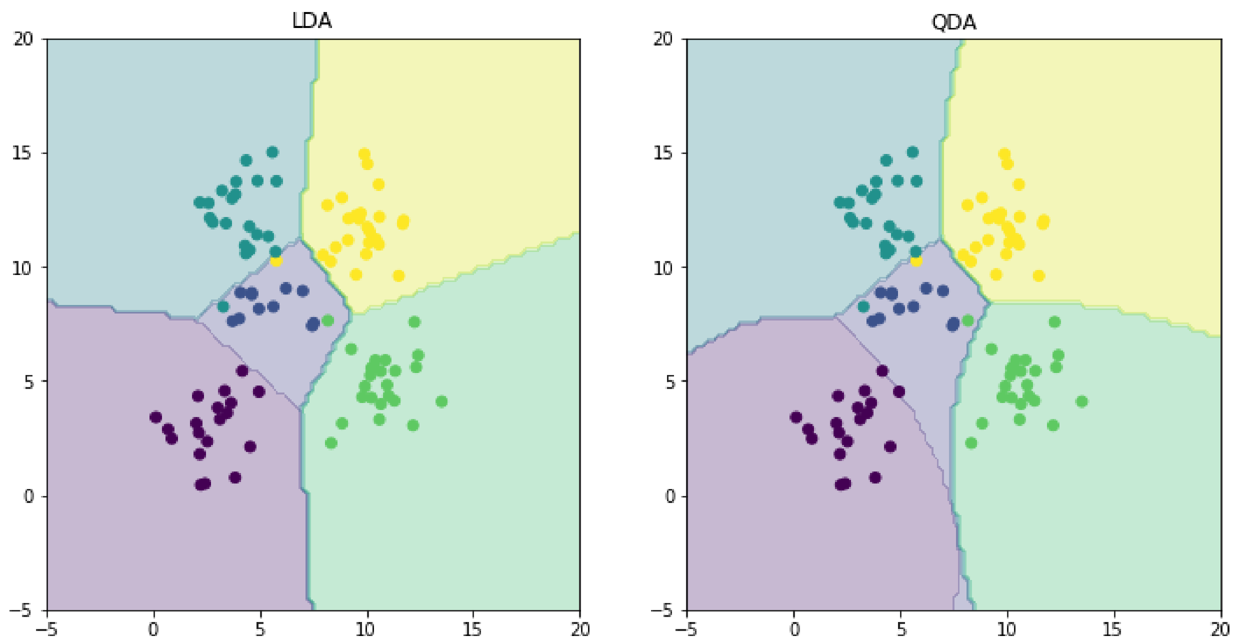
Contents

Problem 1:

Both LDA and QDA methods were used to train sample data and to make predictions on test data.

Obtained accuracy of LDA is .97
Obtained accuracy of QDA is .97

The decision boundary for both LDA and QDA is shown below.



Conclusion:

From the plot above we can see the decision boundary is linear for LDA and non-linear for QDA. This is because in QDA the covariance matrix is different for each class whereas LDA has same covariance matrix for all the classes.

Also, we can observe that both LDA and QDA have made same predictions for the test data. Hence, the accuracy is same for both LDA and QDA. QDA would have provided better results if the training data would have been more.

Problem 2:

Observations:

MSE for train and test with respect to bias and without bias

MSE	With Intercept	Without Intercept
Train	2187.16029	19099.44684
Test	3707.84018	106775.3616

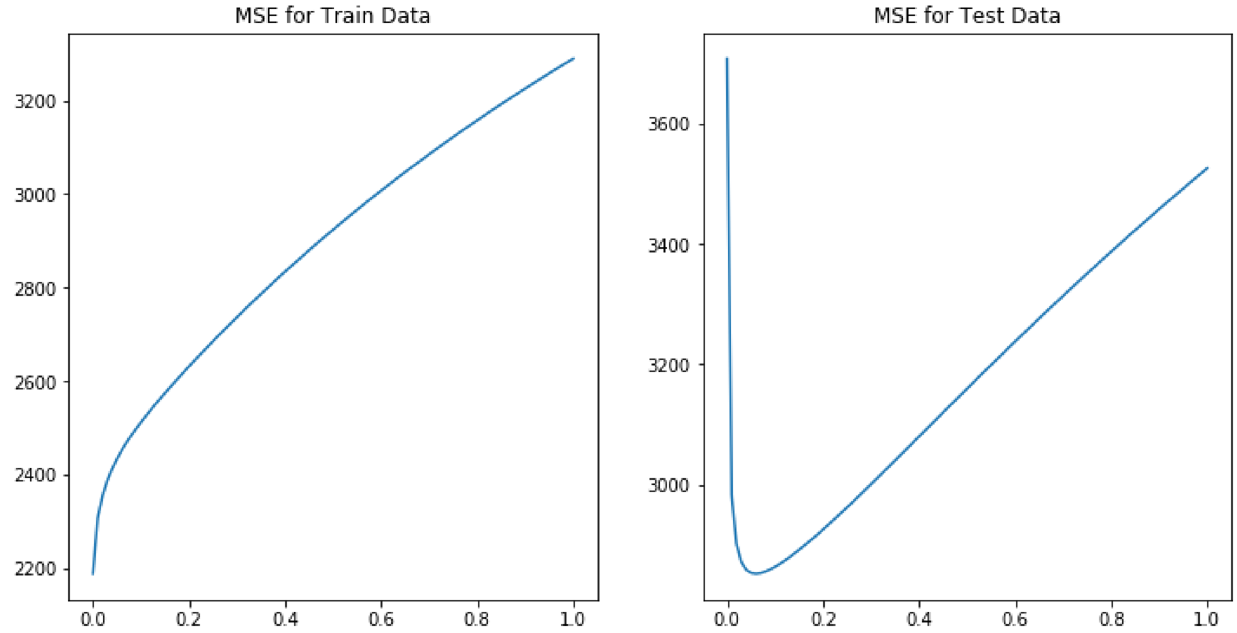
Conclusion:

Without intercept, the line will pass from origin and therefore the error is high. But when bias is added, error is low and we are able to fit the line much better. In both train and test data, MSE is minimum for “with intercept” case.
Hence, “with intercept” case is better.

Problem 3:

Observations:

Plots for train and test data for different lambda values are shown below.



For training data, we can see that as the lambda value increase the MSE increases. This is expected because with increase in lambda value, the model doesn`t learn enough about the data and hence risk underfitting the data.

For test data, we can see that the MSE is minimum with value 2851.33021344 at lambda value is .06. From there on, the MSE increases with increase in lambda.

Hence, the optimal lambda value for this dataset is 0.06.

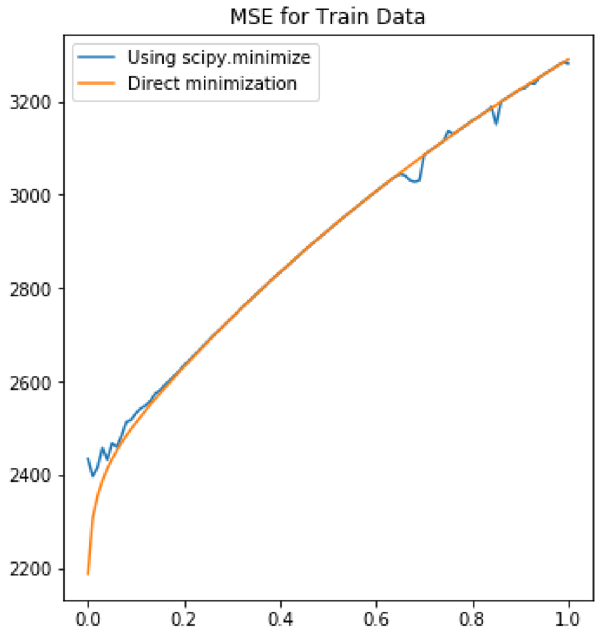
Comparing the weights of OLE and Ridge:

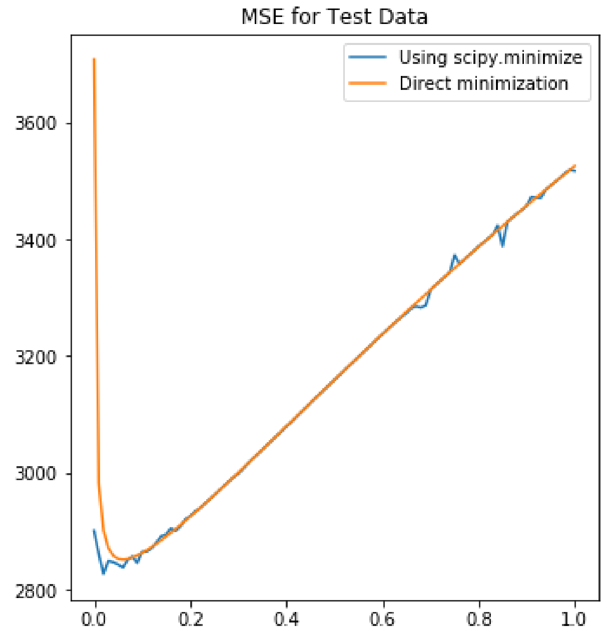
	w_ole	w_ridge
0	-4.121733e+02	150.194724
1	-3.459403e+02	21.701754
2	5.788141e+02	-39.060152
3	5.892438e+01	189.768796
4	-1.358916e+06	131.436318
5	1.194623e+06	12.871668
6	5.070365e+05	-12.647067
7	-1.345870e+03	-111.677408
8	4.477133e+05	99.438088
9	4.779038e+02	203.314471
10	-1.406584e+02	112.568603
11	-9.193403e+02	27.873071
12	-3.959689e+02	55.608089
13	-7.256926e+04	35.949557
14	-8.950937e+04	9.320489
15	-3.237826e+03	-14.822272
16	1.407300e+03	2.083103
17	3.917952e+04	26.192735
18	2.650843e+02	-11.762473
19	5.128435e+02	33.196907
20	2.011581e+02	41.121910
21	6.991405e+01	-1.629232
22	-4.243070e+03	37.817031
23	3.446449e+03	-26.615919

We can see that weights of Ridge at optimal lambda = 0.06 are comparatively low with respect to OLE. This is expected as the weights are penalized in Ridge.

Problem 4:

Observations:





From the plot we can see MSE obtained through gradient descent is less.

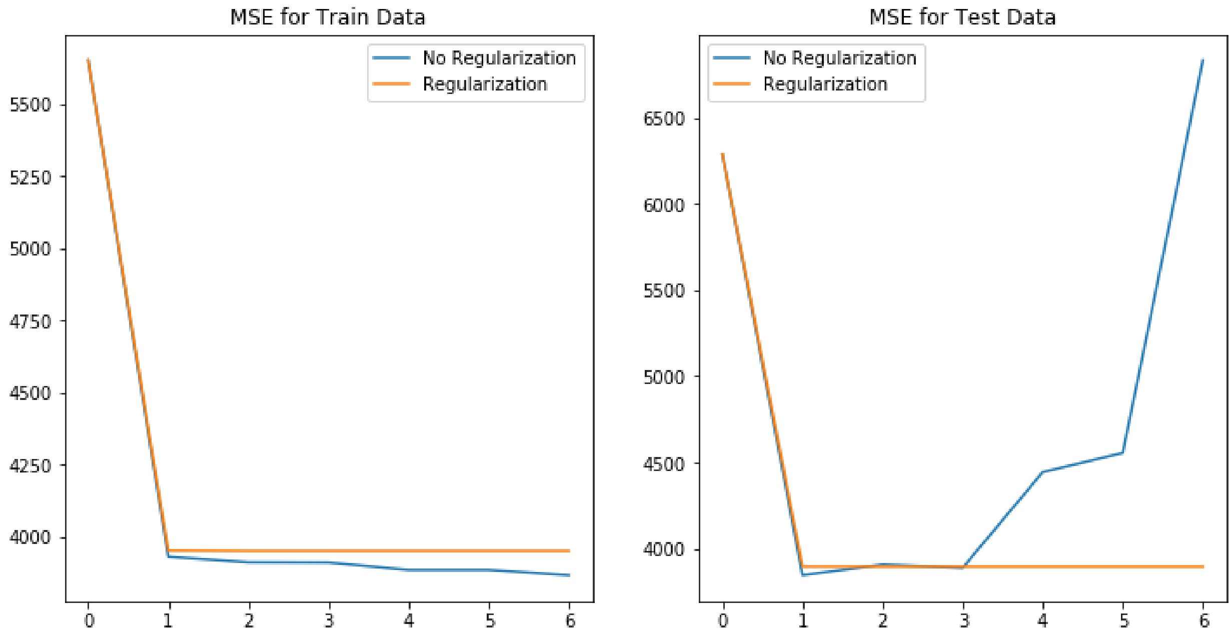
Lambda	With Gradient Descent
0.02	2826.952565

Lambda	Without Gradient Descent
0.06	2851.330213

Thus with Gradient Descent we get better result and the optimal value of lambda in this case is 0.02.

Problem 5:

Observation



From the above plot we can see MSE for train and Test data with regularization and without regularization.

Without regularization, MSE for train data decreases. This is because as p (order of the polynomial) increases, the model tends to overfit the data. Hence, we observe the MSE for test data increasing with increase in p.

But when regularization is used, the effect of overfit is reduced because of the penalization of coefficients. In this case, the behavior of train and test MSE with increase in P almost remains same.

For test data, the MSE is minimum at

- ➔ p = 1 when there is no Regularization
- ➔ p = 4 when there is Regularization (although there isn`t significant change in MSE from p = 1 to p = 6)

Hence optimal value of p = 1 when there is no regularization and optimal value of p = 4 when there is regularization.

Problem 6:

Comparison of various above methods:

The below table shows the comparison of train and test MSE between different methods.

MSE	With Bias	Without Bias	Ridge ($\lambda = 0.06$)	Gradient Descent (Regularization $\lambda = 0.02$)	Non-linear regression (No Regularization, order of polynomial p = 1)	Non-linear regression (Regularization, order of polynomial p = 4)
Train	2187.16029	19909.44684	2451.52849064	2415.24416573	3930.91540732	3950.6823368
Test	3707.84018	106775.3616	2851.33021344	2826.952565	3845.03473017	3895.58266828

The gradient descent based learning method with regularization parameter as lambda = 0.02 has the least MSE for test data when compared to other methods.

Metric to choose Best Setting:

From all the above methods, the best setting should be the one where we achieve minimum MSE on test data without overfitting. The gradient descent based learning method with regularization parameter as lambda = 0.02 agrees with this and hence should be considered as the best setting.